Efficient inference of generalized spatial fusion models with flexible specification

Wang, Craig ; Furrer, Reinhard

Abstract: In spatial statistics, data are often collected at different spatial resolutions. Often, it is of interest to (a) carry out multivariate analysis when variables are sampled at different locations, (b) model data collected at misaligned areas, or (c) unravel common latent factors by jointly modelling point and areal data. In this paper, we establish a linkage between the generalized spatial fusion model framework and the various change-of-support problems, and we outline how the framework can be adapted in these situations. Moreover, we propose an efficient fusion model implementation by exploiting advantages of nearest neighbour Gaussian process and the Stan modelling language. Our simulation shows that the computational efficiency is several times higher in the new implementation compared with original implementation. We illustrate the performance gain in practice using a case study, which models daily precipitation in Switzerland based on rain gauge and radar data.

DOI: https://doi.org/10.1002/sta4.216

Received: 9 November 2018 | Accepted: 26 November 2018
DOI: 10.1002/sta4.216

WILEY

ORIGINAL ARTICLE

Efficient inference of generalized spatial fusion models with flexible specification

Craig Wang[1] | Reinhard Furrer[1,2]

[1]Department of Mathematics, University of Zurich, Zurich, 8057, Switzerland
[2]Department of Computational Science, University of Zurich, Zurich, 8057, Switzerland

Correspondence
Reinhard Furrer, Department of Mathematics, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland.
Email: reinhard.furrer@math.uzh.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 175529

In spatial statistics, data are often collected at different spatial resolutions. Often, it is of interest to (a) carry out multivariate analysis when variables are sampled at different locations, (b) model data collected at misaligned areas, or (c) unravel common latent factors by jointly modelling point and areal data. In this paper, we establish a linkage between the generalized spatial fusion model framework and the various change-of-support problems, and we outline how the framework can be adapted in these situations. Moreover, we propose an efficient fusion model implementation by exploiting advantages of nearest neighbour Gaussian process and the Stan modelling language. Our simulation shows that the computational efficiency is several times higher in the new implementation compared with original implementation. We illustrate the performance gain in practice using a case study, which models daily precipitation in Switzerland based on rain gauge and radar data.

KEYWORDS
change-of-support problem, data fusion, ecological bias, latent spatial process, noncentred parameterization

1 | INTRODUCTION

In modern scientific investigations, research questions are often complex and multifaceted. As a result, researchers spend an increasing amount of effort to collect, merge, and analyze data from multiple sources. The process of integrating and analyzing multiple sources of data with the aim of obtaining more accurate information is often referred to as data fusion. Applications can be found in different research fields, such as bioinformatics (e.g., Lanckriet, De Bie, Cristianini, Jordan, & Noble, 2004) and sensor networks (e.g., Liggins II, Hall, & Llinas, 2017), just to name a few.

When it comes to data fusion using spatial data, the prediction performance can be improved upon models that use only a single data source (Cowles, Yan, & Smith, 2009; Moraga, Cramb, Mengersen, & Pagano, 2017; Shi & Kang, 2017). However, there are several key challenges in implementing spatial fusion models. First of all, heterogeneity often exists among spatial data, which is reflected in terms of different resolutions and different probability distributions. Spatial data may be collected at different resolutions to protect privacy and for practical reasons. In epidemiology, disease data may be collected at residential locations, or they can be aggregated to the district level to maintain privacy. In remote sensing, environmental variables can be collected at monitoring stations, or they can be derived from satellites. As a result of fusing data from different resolutions, spatial fusion models are often required to handle a change-of-support problem (COSP). A COSP may arise in two situations: (a) when studying a single variable at points or areas that are different from where the variable was observed or (b) when jointly modelling multiple spatial variables with different support (Gotway & Young, 2002). Several spatial fusion models have been proposed and applied to link measurements from monitoring stations and gridded computer model outputs for environmental variables. For example, Fuentes and Raftery (2005) proposed a Bayesian melding model to fuse SO_2 measurements from point and areal data. Berrocal, Gelfand, and Holland (2010) presented a downscaling approach to model ground-level measurements and numerical model outputs for ozone data. Sahu, Gelfand, and Holland (2010) used a measurement error model to fuse precipitation data. Other fusion models work with different spatial resolutions. Nguyen, Cressie, and Braver-

man (2012) proposed a spatial random effect model to fuse pixel-based aerosol optical depth data, and Bourgeois et al. (2012) used a hierarchical model to fuse three different areal measurements on weed distribution. However, these fusion models handle only one particular type of COSP: either point-area or area-area. In addition, many of existing spatial fusion models are limited to have Gaussian-distributed response variables. In spatial models assuming exclusively Gaussian distributions is often inappropriate (Diggle, Tawn, & Moyeed, 1998); particularly in spatial fusion models each variable can have a different distribution. For example, Wang, Puhan, and Furrer (2018) fused two lung disease outcome variables, lung function at the individual level and cause-specific mortality at the area level, which are Gaussian-distributed and Poisson-distributed. Few spatial fusion models have been proposed to handle non-Gaussian-distributed response variables. The areal data in Bourgeois et al. (2012) are all Poisson-distributed. Shi and Kang (2017) proposed a general hierarchical model that fuses different pixel-based remote sensing datasets, where the distributions are members of the exponential family. Similarly, Wang et al. (2018) proposed a generalized spatial fusion model framework that handles data from the exponential family of distributions by fusing both point and areal data. Without constraints on resolution and distribution, spatial fusion models can enable a wider range of applications. In this paper, we illustrate that the framework can allow more flexible specifications, such as handling point-point and area-area COSPs.

Another challenge in spatial fusion models and geostatistical analysis in general is the computation time. The number of floating point operations (flops) required for evaluating the exact log-likelihood of a Gaussian process with n locations is $O(n^3)$. Several solutions exist in spatial statistics to reduce computation time of such an analysis, such as low rank (Banerjee, Gelfand, Finley, & Sang, 2008; Cressie & Johannesson, 2007; Stein, 2008) and sparse (Datta, Banerjee, Finley, & Gelfand, 2016; Furrer, Genton, & Nychka, 2006; Rue, Martino, & Chopin, 2009) methods. Some of them are adapted to existing spatial fusion models. For example, Nguyen et al. (2012) and Shi and Kang (2017) adapted the spatial basis function approach from fixed rank kriging (Cressie & Johannesson, 2007), and the latter used empirical Bayesian estimation. Moraga et al. (2017) used integrated nested Laplace approximations based on Rue et al. (2009). Wang et al. (2018) adapted an implementation of the nearest neighbour Gaussian process (NNGP) (Datta et al., 2016; Zhang, Datta, & Banerjee, 2018) in the Stan modelling language.

In this paper, we also work with Stan, a computational platform for Bayesian inference (Stan Development Team, 2017). It uses the No-U-Turn sampler (Homan & Gelman, 2014), an improved version of Hamiltonian Monte Carlo (HMC), to obtain posterior samples for Bayesian hierarchical models. HMC enables the chains to converge faster than the Gibbs sampler and Metropolis–Hastings algorithm, especially in high dimensional target distributions and for complex hierarchical structures. The algorithm assigns a momentum to each parameter and updates this momentum based on the gradient of log-posterior density at each HMC iteration. The parameters are then updated based on the momentum. Gradient evaluation is known to be a bottleneck in Stan (Stan Development Team, 2017). Nevertheless, it is an accessible and computationally efficient tool for many applications.

In this paper, we propose an alternative implementation of the framework using Stan, which improves the computational speed several fold. The new implementation utilizes noncentred parameterization on the NNGP components and vectorization of log-likelihood evaluation in Stan. The rest of this paper is organized as follows. In Section 2, we formulate the flexible specification of the generalized spatial fusion model framework. In Section 3, we introduce an efficient implementation. In Section 4, we conduct a simulation study to evaluate the performance of our new implementation. Afterwards, we demonstrate the performance gain in practice using a case study on daily precipitation in Section 5. We end with a discussion in Section 6.

2 | GENERALIZED SPATIAL FUSION MODEL FRAMEWORK WITH FLEXIBLE SPECIFICATION

2.1 | Model formulation

We let $\boldsymbol{Y}_j$ denote the jth response variable whose conditional distribution is a member of the exponential family. It is observed either at n_j sites $\boldsymbol{s} \in D \subseteq \mathcal{R}^2$ to have $\boldsymbol{Y}_j = \boldsymbol{Y}_{\mathcal{S}_j} \equiv \left(Y_j(\boldsymbol{s}_{j1}), Y_j(\boldsymbol{s}_{j2}), \ldots, Y_j(\boldsymbol{s}_{jn_j})\right)^T$ with $\mathcal{S}_j = \{\boldsymbol{s}_{j1}, \boldsymbol{s}_{j2}, \ldots, \boldsymbol{s}_{jn_j}\}$ or at n_j areas $\boldsymbol{a} \subset D$ to have $\boldsymbol{Y}_j = \boldsymbol{Y}_{\mathcal{A}_j} \equiv \left(Y_j(\boldsymbol{a}_{j1}), Y_j(\boldsymbol{a}_{j2}), \ldots, Y_j(\boldsymbol{a}_{jn_j})\right)^T$ with $\mathcal{A}_j = \{\boldsymbol{a}_{j1}, \boldsymbol{a}_{j2}, \ldots, \boldsymbol{a}_{jn_j}\}$. Without loss of generality, we consider modelling two response variables with $j = 1, 2$. We assume a zero-mean latent Gaussian process (GP) $w(\boldsymbol{u})$ with covariance function $C(\cdot, \cdot; \boldsymbol{\theta})$ that is associated with both response variables, that is, $w(\boldsymbol{u}) \sim GP(0, C(\cdot, \cdot; \boldsymbol{\theta}))$. Depending on the observed spatial resolutions, we denote the latent process at corresponding locations of the jth response to be $\boldsymbol{w}_j$, with either $\boldsymbol{w}_j = \boldsymbol{w}_{\mathcal{S}_j} \equiv \left(w(\boldsymbol{s}_{j1}), w(\boldsymbol{s}_{j2}), \ldots, w(\boldsymbol{s}_{jn_j})\right)^T$ or $\boldsymbol{w}_j = \boldsymbol{w}_{\mathcal{A}_j} \equiv \left(w(\boldsymbol{a}_{j1}), w(\boldsymbol{a}_{j2}), \ldots, w(\boldsymbol{a}_{jn_j})\right)^T$, where $w(\boldsymbol{a}_{ji})$ is the aggregated process for area $\boldsymbol{a}_{ji}$ with $i = 1, 2, \ldots, n_j$.

Under a generalized framework, the spatial fusion model can be formulated as

$$g_j\left(\mathbb{E}\left[\boldsymbol{Y}_j \mid \boldsymbol{\beta}_j, \boldsymbol{w}_j\right]\right) = \boldsymbol{X}_j^T \boldsymbol{\beta}_j + \boldsymbol{w}_j, \tag{1}$$

where $g_j(\cdot)$ is a suitable link function that corresponds to the conditional distribution of the response variable, for example, The log link for Poisson distribution and the logit link for binomial distribution, $\boldsymbol{X}_j$ is a matrix of geo-referenced covariates for the jth response variable, $\boldsymbol{\beta}_j$ is a vector of the corresponding coefficients, and $\boldsymbol{w}_j$ has the same spatial resolution as $\boldsymbol{Y}_j$.

The original formulation of the framework was introduced in Wang et al. (2018) with $\boldsymbol{Y}_1 = \boldsymbol{Y}_{\mathcal{S}_1}, \boldsymbol{w}_1 = \boldsymbol{w}_{\mathcal{S}_1}, \boldsymbol{Y}_2 = \boldsymbol{Y}_{\mathcal{A}_2}$, and $\boldsymbol{w}_2 = \boldsymbol{w}_{\mathcal{A}_2}$, that is, to jointly model point and areal data. However, the flexibility of this framework extends beyond those choices. The framework additionally allows for both responses to be at the point level or at the area level, which we outline in Section 2.2.

2.2 | Change-of-support problems

COSP need to be addressed as long as $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ have different spatial support, which one can expect to occur often because fusion tasks normally involve different data sources. Consider the simplest case with two point-referenced response variables that are Gaussian-distributed, namely, $\boldsymbol{Y}_{\mathcal{S}_1}$ and $\boldsymbol{Y}_{\mathcal{S}_2}$ with $\mathcal{S}_1 \neq \mathcal{S}_2$. Because we assume there is a common latent GP for both responses, we let the set of locations in the latent process $\mathcal{U}$ to comprise the observed locations from each response, that is, $\mathcal{U} = \mathcal{S}_1 \cup \mathcal{S}_2$. This scenario can be formulated as

$$\begin{aligned} \boldsymbol{Y}_{\mathcal{S}_1} \mid \boldsymbol{\beta}_1, \boldsymbol{w}_{\mathcal{S}_1}, \tau_1^2 &\sim \mathcal{N}\left(\boldsymbol{X}_1^T\boldsymbol{\beta}_1 + \boldsymbol{w}_{\mathcal{S}_1}, \tau_1^2\boldsymbol{I}\right), \\ \boldsymbol{Y}_{\mathcal{S}_2} \mid \boldsymbol{\beta}_2, \boldsymbol{w}_{\mathcal{S}_2}, \tau_2^2 &\sim \mathcal{N}\left(\boldsymbol{X}_2^T\boldsymbol{\beta}_2 + \boldsymbol{w}_{\mathcal{S}_2}, \tau_2^2\boldsymbol{I}\right), \\ \boldsymbol{w}_{\mathcal{U}} &\sim GP(0, C(\cdot,\cdot;\boldsymbol{\theta})), \end{aligned} \tag{2}$$

where τ_1^2 and τ_2^2 are variance terms. The Gaussian distributions can be replaced by any other members of the exponential family. When doing so, the variance terms are no longer necessary if the replaced distributions already have a stochastic component (Banerjee, Carlin,& Gelfand, 2014, p. 138).

When one of the response variables is at the point level and the other is at the area level, namely, $\boldsymbol{Y}_{\mathcal{S}_1}$ and $\boldsymbol{Y}_{\mathcal{A}_2}$, we obtain the formulation introduced by Wang et al. (2018). The aggregation of a latent process to $\boldsymbol{w}_{\mathcal{A}_2}$ requires stochastic integrals, which can be approximated using sampling points. If the conditional distribution of $\boldsymbol{Y}_{\mathcal{A}_2}$ remains Gaussian, then

$$w(\boldsymbol{a}_{2i}) = |\boldsymbol{a}_{2i}|^{-1}\int_{\boldsymbol{u}\in\boldsymbol{a}_{2i}} w(\boldsymbol{u})d\boldsymbol{u} \approx \frac{1}{L}\sum_{k=1,\boldsymbol{s}'_{2ik}\in\boldsymbol{a}_{2i}}^{L} w(\boldsymbol{s}'_{2ik}), \tag{3}$$

where $\boldsymbol{s}'_{2ik}$ is the kth sampling point within area $\boldsymbol{a}_{2i}$, L is the number of sampling points in each area, and we denote the set of all sampling points to be $\mathcal{S}'_2$. This approximation via sampling points was also used by Berrocal et al. (2010), Fuentes and Raftery (2005), and Liu, Le, and Zidek (2011). It has been shown that a small L serves a reasonable trade-off between computational efficiency and model accuracy. If the conditional distribution is non-Gaussian, ecological bias (Greenland, 1992) can occur as a result of aggregating under a non-linear link function. The association between the response variable and the aggregated latent process in (3) is different from that between the response and the original latent process. Such bias can be addressed under the sampling points approach. For example, if we assume $\boldsymbol{Y}_{\mathcal{A}_2}$ is now Poisson-distributed and is modelled with a log link function, then we have the following hierarchical model:

$$\begin{aligned} \boldsymbol{Y}_{\mathcal{S}_1} \mid \boldsymbol{\beta}_1, \boldsymbol{w}_{\mathcal{S}_1}, \tau_1^2 &\sim \mathcal{N}\left(\boldsymbol{X}_1^T\boldsymbol{\beta}_1 + \boldsymbol{w}_{\mathcal{S}_1}, \tau_1^2\boldsymbol{I}\right), \\ \boldsymbol{Y}_{\mathcal{A}_2} \mid \boldsymbol{\beta}_2, \boldsymbol{w}_{\mathcal{A}_2} &\sim \text{Poisson}\left(\exp(\boldsymbol{X}_2^T\boldsymbol{\beta}_2 + \boldsymbol{w}_{\mathcal{A}_2})\right), \\ \boldsymbol{w}_{\mathcal{U}} &\sim GP(0, C(\cdot,\cdot;\boldsymbol{\theta})), \end{aligned} \tag{4}$$

where the approximation of stochastic integrals differs from (3) to address ecological bias, being

$$w(\boldsymbol{a}_{2i}) = \log\left(|\boldsymbol{a}_{2i}|^{-1}\int_{\boldsymbol{u}\in\boldsymbol{a}_{2i}} \exp(w(\boldsymbol{u}))d\boldsymbol{u}\right) \approx \log\left(\frac{1}{L}\sum_{k=1,\boldsymbol{s}'_{2ik}\in\boldsymbol{a}_{2i}}^{L} \exp(w(\boldsymbol{s}'_{2ik}))\right). \tag{5}$$

The latent process is now modelled at both observed locations and sampling points, with $\mathcal{U} = \mathcal{S}_1 \cup \mathcal{S}'_2$.

The final scenario is when both response variables are observed at the area level but with different alignment, that is, having $\boldsymbol{Y}_{\mathcal{A}_1}$ and $\boldsymbol{Y}_{\mathcal{A}_2}$. This is also called the modifiable areal unit problem (Gotway & Young, 2002). As an example, if both response variables are Poisson-distributed, the model can be formulated as

$$\begin{aligned} \boldsymbol{Y}_{\mathcal{A}_1} \mid \boldsymbol{\beta}_1, \boldsymbol{w}_{\mathcal{A}_1} &\sim \text{Poisson}\left(\exp(\boldsymbol{X}_1^T\boldsymbol{\beta}_1 + \boldsymbol{w}_{\mathcal{A}_1})\right), \\ \boldsymbol{Y}_{\mathcal{A}_2} \mid \boldsymbol{\beta}_2, \boldsymbol{w}_{\mathcal{A}_2} &\sim \text{Poisson}\left(\exp(\boldsymbol{X}_2^T\boldsymbol{\beta}_2 + \boldsymbol{w}_{\mathcal{A}_2})\right), \\ \boldsymbol{w}_{\mathcal{U}} &\sim GP(0, C(\cdot,\cdot;\boldsymbol{\theta})), \end{aligned} \tag{6}$$

where sampling points are required for both $\mathcal{A}_1$ and $\mathcal{A}_2$, which we denote as $\mathcal{S}'_1$ and $\mathcal{S}'_2$. For computational efficiency, $\mathcal{S}'_1$ and $\mathcal{S}'_2$ can have an overlapping set of sampling points and $\mathcal{U} = \mathcal{S}'_1 \cup \mathcal{S}'_2$.

3 | EFFICIENT IMPLEMENTATION

The computational cost is high when a Gaussian process is fitted in a Bayesian hierarchical model. Convenient techniques in Markov chain Monte Carlo-based Bayesian inference, such as using conjugate priors and marginalization, are only available in geostatistical models with Gaussian-distributed response variables. The projection-based method (Guan & Haran, 2018) for non-Gaussian spatial data has shown promising performance, but it is not directly applicable to fusion models. Recently, Datta et al. (2016) proposed NNGP as a sparse approximation to a full GP. The prediction performance is indistinguishable between the two, whereas NNGP only uses a fraction of the computation time. In the previous implementation of the spatial fusion model, Wang et al. (2018) utilized NNGP in modelling the latent process and implemented fusion

models in Stan (Stan Development Team, 2017). Such implementation allowed greater computational efficiency compared with using a full GP. In this section, we propose an efficient implementation of the fusion model framework to further reduce computation time. It utilizes vectorization in Stan and applies noncentred parameterization to both location and scale parameters of the latent NNGP.

3.1 | Vectorization

Although using NNGP increase the speed of log-likelihood computation compared with a full GP, its gradient evaluation in Stan can be slow due to non-standard and nonvectorized probability functions in the sampling statements. In order to mitigate this drawback, we reimplement the NNGP log-likelihood using vectorized, conditionally independent Gaussian log-likelihoods.

Following Datta et al. (2016), the construction of NNGP likelihood for $\boldsymbol{w}_{\mathcal{U}^*}$ requires a fixed set of reference locations. We select the union of observed and sampling locations $\mathcal{U}^*$ in the fusion framework as the reference set, which can reduce computational cost by preventing additional reference locations from appearing in the model. The NNGP log-likelihood for $\boldsymbol{w}_{\mathcal{U}^*}$ can be rewritten as

$$\tilde{p}(\boldsymbol{w}_{\mathcal{U}^*}) = \sum_{i=1}^{n_s+n_aL} log\left(\mathcal{N}\left(w(\boldsymbol{u}_i) \mid C_{\boldsymbol{u}_i,N(\boldsymbol{u}_i)}C^{-1}_{N(\boldsymbol{u}_i)}w_{N(\boldsymbol{u}_i)}, C_{\boldsymbol{u}_i,\boldsymbol{u}_i} - C_{\boldsymbol{u}_i,N(\boldsymbol{u}_i)}C^{-1}_{N(\boldsymbol{u}_i)}C^{\top}_{\boldsymbol{u}_i,N(\boldsymbol{u}_i)}\right)\right), \tag{7}$$

where $N(\boldsymbol{u}_i)$ is the set of $\max(i-1, m)$ nearest neighbours from $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{i-1}\}$ for location $\boldsymbol{u}_i$, $C_{\boldsymbol{u}_i,N(\boldsymbol{u}_i)}$ is the cross-covariance matrix between the latent process $w(\boldsymbol{u}_i)$ and its neighbours $N(\boldsymbol{u}_i)$, $C_{N(\boldsymbol{u}_i)}$ is the covariance matrix of $\boldsymbol{w}_{N(\boldsymbol{u}_i)}$, and $C(\boldsymbol{u}_i, \boldsymbol{u}_i)$ is the variance of $w(\boldsymbol{u}_i)$. The dimensions of covariance and cross-covariance matrices depend on the index i. For example, $C_{\boldsymbol{u}_i,N(\boldsymbol{u}_i)}$ is a $1 \times (i-1)$ matrix for $i \leq m$ and is a $1 \times m$ matrix for $i > m$.

As it can be seen in (7), the mean of the latent process $w(\boldsymbol{u}_i)$ for $i > 1$ depends on $w_{N(\boldsymbol{u}_i)}$, which consists of the latent process at locations with smaller indices. When implementing such log-likelihood, one may choose to update each $w(\boldsymbol{u}_i)$ sequentially in a Gibbs sampling style. However, this can lead to slow convergence for a large number of locations. Stan modelling language allows user-defined probability functions. An acceptable approach is to implement (7) directly in Stan where it then does block updates on $\boldsymbol{w}_{\mathcal{U}^*}$. This implementation was used by Zhang et al. (2018) and adopted for the fusion model in Wang et al. (2018). However, this approach causes slow gradient evaluation due to custom NNGP log-likelihood. In order to use standard probability functions and to vectorize the sampling statements, we first compute the conditional means and variances according to (7) for $w(\boldsymbol{u}_i)$ based on the nearest neighbours $\boldsymbol{w}_{N(\boldsymbol{u}_i)}$, then draw the current iteration of $\boldsymbol{w}_{\mathcal{U}^*}$ from $n_s + n_aL$ number of conditionally independent normal distributions with the computed conditional mean and variance. This allows block-updating on $\boldsymbol{w}_{\mathcal{U}^*}$ and saves gradient evaluation time.

3.2 | Noncentred parameterization

Because the latent process cannot be marginalized in the generalized fusion models, $\boldsymbol{w}_{\mathcal{U}^*}$ is treated as parameters that must be sampled in Bayesian hierarchical models. Although Stan has the advantage of faster mixing by using the No-U-Turn sampler, sampling-correlated parameters may still be slow in a high-dimensional parameter space, such as with Monte Carlo sampling algorithms like Random Walk Metropolis and Gibbs sampler (Betancourt & Girolami, 2013). Hence, we apply noncentred reparameterization (Papaspiliopoulos, Roberts, & Sköld, 2007) to both location and scale parameters of the vectorized latent process. This implies that the centred parameterization

$$w(\boldsymbol{u}_i) \mid w_{N(\boldsymbol{u}_i)}, \boldsymbol{\theta} \sim \mathcal{N}\left(C_{\boldsymbol{u}_i,N(\boldsymbol{u}_i)}C^{-1}_{N(\boldsymbol{u}_i)}w_{N(\boldsymbol{u}_i)}, C_{\boldsymbol{u}_i,\boldsymbol{u}_i} - C_{\boldsymbol{u}_i,N(\boldsymbol{u}_i)}C^{-1}_{N(\boldsymbol{u}_i)}C^{\top}_{\boldsymbol{u}_i,N(\boldsymbol{u}_i)}\right), \tag{8}$$

becomes

$$w(\boldsymbol{u}_i) \mid w_{N(\boldsymbol{u}_i)}, \boldsymbol{\theta} = C_{\boldsymbol{u}_i,N(\boldsymbol{u}_i)}C^{-1}_{N(\boldsymbol{u}_i)}w_{N(\boldsymbol{u}_i)} + \tilde{w}(\boldsymbol{u}_i)\left(C_{\boldsymbol{u}_i,\boldsymbol{u}_i} - C_{\boldsymbol{u}_i,N(\boldsymbol{u}_i)}C^{-1}_{N(\boldsymbol{u}_i)}C^{\top}_{\boldsymbol{u}_i,N(\boldsymbol{u}_i)}\right)^{1/2} \text{ with } \tilde{w}(\boldsymbol{u}_i) \sim \mathcal{N}(0,1), \tag{9}$$

after reparameterization. This removes the correlation in the actively sampled latent process parameters, leaving us to sample from standard normal distributions.

4 | SIMULATION STUDY

In this section, we conduct a simulation study to compare the performance of different fusion model implementations. All simulation results are obtained in R version 3.3.3 (R Core Team, 2018) with rstan version 2.16.2 Carpenter et al. (2017), on a Linux server with 256 GB of RAM and two Intel Xeon 6-core 2.5 GHz processors. The Stan model code for all implementations is provided in the Supporting Information available online.

4.1 | Simulation set-up

We are interested in modelling a latent spatial process within a $[0, 4000] \times [0, 4000]$ domain. First, we generate a zero-mean GP $w(\boldsymbol{u})$ on a fine grid with covariance matrix $C(\cdot, \cdot; \theta)$. We then subsample n_s locations from the fine grid to obtain $\boldsymbol{w}_{S_1}$ at observed locations. For areal observations, we divide the domain into n_a number of Voronoi cells and compute aggregated $\boldsymbol{w}_{\mathcal{A}_2}$. Afterwards, we generate a covariate for each location and a

covariate for each area from standard normal distributions. By adding intercepts, we obtain $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. The response variables are then generated according to

$$\begin{aligned} \boldsymbol{Y}_{\mathcal{S}_1} \mid \boldsymbol{\beta}_1, \boldsymbol{w}_{\mathcal{S}_1}, \tau_1^2 &\sim \mathcal{N}\left(\boldsymbol{X}_1^T\boldsymbol{\beta}_1 + \boldsymbol{w}_{\mathcal{S}_1}, \tau_1^2\boldsymbol{I}\right), \\ \boldsymbol{Y}_{\mathcal{A}_2} \mid \boldsymbol{\beta}_2, \boldsymbol{w}_{\mathcal{A}_2} &\sim \text{Poisson}\left(\exp(\boldsymbol{X}_2^T\boldsymbol{\beta}_2 + \boldsymbol{w}_{\mathcal{A}_2})\right). \end{aligned} \tag{10}$$

Because the areal data are conditionally Poisson-distributed, (5) is used to aggregate the latent spatial process to compute $w(\boldsymbol{a})$ for each area. In the simulation, we use an exponential covariance function from the Matérn family, that is, $C(\boldsymbol{s}_i, \boldsymbol{s}_j; \sigma^2, \phi) = \sigma^2 \exp(-||\boldsymbol{s}_i - \boldsymbol{s}_j||/\phi)$, where $||\boldsymbol{s}_i - \boldsymbol{s}_j||$ is the Euclidean distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, σ^2 is the partial sill that represents spatial covariance, and ϕ is the decay parameter that controls the spatial range.

We simulate a total of 40 scenarios with a combination of different sample sizes and parameters, where $n_s = \{50, 200, 500\}$, $n_a = \{25, 75, 200\}$, and $(\sigma^2, \tau^2, \phi) = \{(0.1, 1.0, 300), (0.5, 0.5, 300)\}$. The coefficients are $\boldsymbol{\beta}_1 = (1, 5)^T, \boldsymbol{\beta}_2 = (1, 2)^T$. We consider the following three model specifications: (a) the original fusion model implementation from Wang et al. (2018); (b) the fusion model with vectorization only (denoted as fusion.vec); and (c) the fusion model with vectorization and noncentred parameterization (denoted as fusion.ncp). For all models, the intercepts and coefficients are assigned with independent $N(0, 5^2)$ priors. The variance parameters σ^2 and τ^2 are assigned with an inverse Gamma prior $IG(2, 1)$, which has a mean of one and undefined variance. For spatial decay, a normal prior $N(300, 100^2)$ truncated at zero is assigned.

We use $m = 5$ nearest neighbours and $L = 5$ sampling points randomly selected within each area. We run four chains of 2,000 iterations with 1,000 warm-up samples, without thinning for each model. Multiple chain convergence is checked with potential scale reduction factors (Brooks & Gelman, 1998). We compute the relative effective sample size (ESS; Gelman et al., 2013) per minute separately for $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \tau^2, \sigma^2, \phi\}$ and for the latent process using the original implementation as the baseline for comparison.

TABLE 1 Simulation results with different combinations of sample sizes for both scenarios

			Scenario 1: $\sigma^2 = 0.5, \tau^2 = 0.5$					Scenario 2: $\sigma^2 = 0.1, \tau^2 = 1$				
n_1	n_2	Method	RMSPE	Cov	Time	reESS	reESSw	RMSPE	Cov	Time	reESS	reESSw
25	50	fusion	0.644	91.8	2.5	1	1	0.302	100	2.4	1	1
		fusion.vec	0.644	92.2	1.7	1.58	1.55	0.302	100	1.6	1.5	1.5
		fusion.ncp	0.644	92.2	1.7	2.11	1.79	0.303	100	1.5	2.61	1.82
	200	fusion	0.537	95.9	5.9	1	1	0.310	98.6	6.5	1	1
		fusion.vec	0.537	95.6	3.8	1.38	1.15	0.309	98.6	5.0	1.31	1.25
		fusion.ncp	0.537	95.9	4.6	2.54	1.61	0.310	98.6	3.0	3.61	2.27
	500	fusion	0.489	97.4	11.5	1	1	0.290	99.1	18.2	1	1
		fusion.vec	0.490	97.7	8.2	1.31	1.43	0.290	99.2	12.2	1.36	1.54
		fusion.ncp	0.490	97.6	9.0	1.91	1.33	0.290	99.2	5.0	5.8	4.37
75	50	fusion	0.624	95.1	9.2	1	1	0.295	99.1	9.8	1	1
		fusion.vec	0.615	95.2	6.1	1.64	1.82	0.296	99.0	6.0	1.6	1.59
		fusion.ncp	0.616	95.5	6.9	2.66	1.84	0.295	98.9	4.6	2.82	2.27
	200	fusion	0.534	93.8	11.8	1	1	0.274	99.1	18.8	1	1
		fusion.vec	0.534	93.8	8.1	1.55	1.57	0.276	99.1	11.9	1.04	1.56
		fusion.ncp	0.533	93.8	9.1	3.22	1.48	0.274	99.1	6.7	3.29	3.11
	500	fusion	0.519	91.4	21.1	1	1	0.270	99.3	47.1	1	1
		fusion.vec	0.518	91.6	14.1	1.41	1.34	0.270	99.2	29.6	1.54	1.56
		fusion.ncp	0.517	91.9	14.1	3.72	1.69	0.270	99.1	31.3	2.89	2.02
200	50	fusion	0.623	94.3	41.8	1	1	0.272	99.8	51.1	1	1
		fusion.vec	0.625	94.4	24.9	1.67	1.18	0.272	99.6	32.3	1.53	1.6
		fusion.ncp	0.623	94.2	29.0	2.12	3.56	0.272	99.7	26.2	3.7	2.02
	200	fusion	0.504	96.6	65.2	1	1	0.244	99.5	52.9	1	1
		fusion.vec	0.504	96.6	39.0	1.61	1.2	0.245	99.3	32.7	1.76	1.66
		fusion.ncp	0.504	96.6	42.0	4.13	2.06	0.244	99.4	22.6	4.29	2.45
	500	fusion	0.456	96.9	63.1	1	1	0.278	97.6	100.4	1	1
		fusion.vec	0.456	96.8	41.8	1.39	1.66	0.273	97.4	70.0	1.46	1.65
		fusion.ncp	0.456	96.8	35.4	2.97	2.44	0.275	96.4	62.1	3.67	2.18

Note. The number of point observations and area observations are denoted by n_1 and n_2, respectively. RMSPE is the root mean squared prediction error and Cov is the coverage probability of 95% credible interval for the latent process at prediction locations. Time represents the computation time in minutes for running four chains in parallel. Finally, reESS and reESSw represent the relative time-standardized effective sample sizes for $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \tau^2, \sigma^2, \phi\}$ and for the latent process and are used to measure computational efficiency relative to the original fusion implementation.

We chose an additional 1,600 sites under each scenario to evaluate predictive performance. The prediction sites are located at the centres of a 40 × 40 grid that uniformly covers the sampling domain. To compare the models' prediction performance, we compute the coverage probabilities of 95% credible intervals and root mean squared prediction errors (RMSPE).

4.2 | Simulation results

Table 1 summarizes the simulation results comparing different fusion model implementations. There is negligible difference among them in terms of their RMSPE and coverage probability for the predicted latent process. Generally, the RMSPE decreases as the sample size increases. Scenario 1 has a higher RMSPE than Scenario 2, because the latent process has a higher variability in terms of partial sill. The reduction in computation time due to vectorization is reflected by comparing the first two methods, where the time required for gradient evaluation is immediately reduced. The methods' efficiency in terms of ESS per minute differs significantly. When considering the original fusion model implementation as the baseline, the fusion.vec improves the efficiency in terms of relative ESS by avoiding the gradient evaluation of custom NNGP likelihood and vectorization. However, drawing high dimensional correlated parameters can slow down the sampling process. When applying noncentred parameterization in fusion.ncp, the efficiency is further improved.

5 | CASE STUDY

In this section, we apply a generalized spatial fusion model to map daily precipitation data in Switzerland and compare the performance of different fusion model implementations. We select a particular day on June 3, 2017, where the majority of Switzerland experienced some rainfall. The precipitation data come from two sources, namely, rain gauge measurements and radar images. Rain gauge data for hourly total precipitation in millimetres (mm) were collected from 536 monitoring stations distributed across Switzerland. Radar images were sampled at 5-min time intervals at a spatial resolution of 1 km × 1 km from the MeteoSwiss fourth generation weather radar network, resulting in 41,278 grid cells. They have been been calibrated to various external information (MeteoSwiss, 2018). The advantage of station measurements is that they have precise geo-coordinates that allow for geostatistical modelling. However, modelling precipitation based only on station measurements cannot well characterize the spatial structure of precipitation as a result of their limited spatial coverage. Therefore, combining rain gauge measurements and radar images can improve the resulting map (e.g., Fuentes, Reich, & Lee, 2008; Sideris, Gabella, Erdin, & Germann, 2014).

In the spatial fusion model, we include elevation as a covariate for the precipitation. The elevation data have a resolution of 200 m, which is resampled from a digital height model of 25-m resolution. For the purpose of this case study, we aggregate the grid cells from radar images to 20 km × 20 km cells. For cells with high variations in elevation, as measured by the coefficient of variation, we further divide the cell into four smaller cells to result 255 cells in the final analysis. Both data sources are temporally aggregated to daily precipitation. Figure 1 shows the daily precipitation at both the monitoring stations and the aggregated radar grids. The fusion model in this case study is similar to a zero-inflated log-Gaussian process proposed by Fuentes et al. (2008); however, we model the latent process as a GP instead of a Gaussian Markov random field. The model contains two stages. The first stage models the amount of rain $\boldsymbol{Y}_{\mathcal{S}_1}$ and $\boldsymbol{Y}_{\mathcal{A}_2}$, which are used to denote log-transformed precipitation measured by rain gauges and aggregated radar images, respectively. The second stage models the occurrence of precipitation $I_{\mathcal{S}_1}$ and $I_{\mathcal{A}_2}$, which are indicator variables on whether the gauge or radar measures some precipitation. The model can be written as

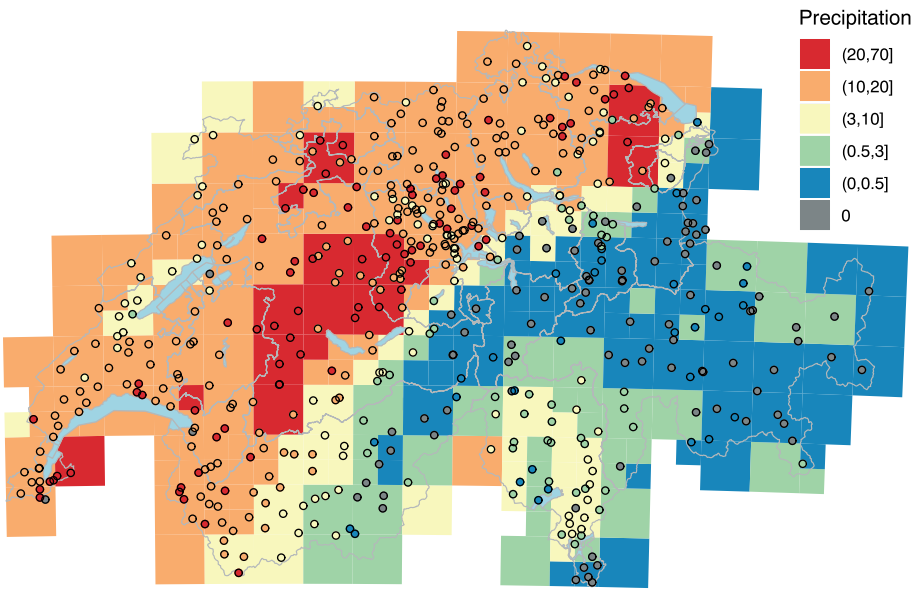

FIGURE 1 Map of total precipitation (mm) on June 3, 2017, in Switzerland in the log scale. Filled circles represent rain gauge measurements and background colour grid represents radar measurements. Grey circles indicates zero rain gauge measurements

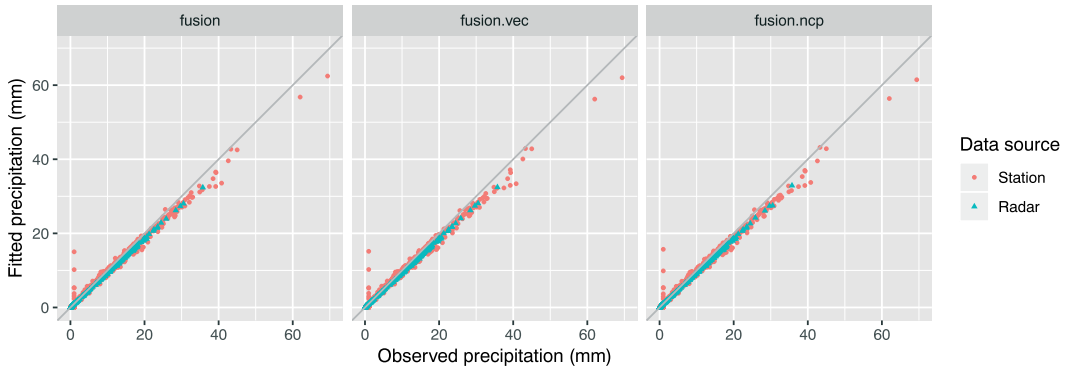

FIGURE 2 Plots of fitted versus observed data for different model implementations. The *y*-axis shows posterior medians of the fitted precipitation after exponential transformation from the log scale

TABLE 2 Comparison of model performance for the case study

Model	RMSE	Time	reESS	reESSw
fusion	1.315	33.5	1	1
fusion.vec	1.325	22.7	1.78	6.59
fusion.ncp	1.335	9.3	1.20	81.74

Note. RMSE is the root mean squared error for the fitted precipitations; reESS and reESSw represent the relative time-standardized effective sample sizes for $\{\boldsymbol{\beta}_1, \tau^2, \sigma^2, \phi\}$ and for the latent process and are used to measure computational efficiency. Finally, time represents the computation time in hours for running four chains in parallel.

$$
\begin{aligned}
\boldsymbol{Y}_{\mathcal{S}_1} \mid \boldsymbol{\beta}_1, \boldsymbol{w}_{\mathcal{S}_1}, \tau_1^2 &\sim \mathcal{N}(\beta_{1,0} + \beta_{1,1} \times \text{elevation}_{\mathcal{S}_1} + \boldsymbol{w}_{\mathcal{S}_1}, \tau_1^2 \boldsymbol{I}),\\
\boldsymbol{Y}_{\mathcal{A}_2} \mid \boldsymbol{\beta}_2, \boldsymbol{w}_{\mathcal{A}_2}, \tau_2^2 &\sim \mathcal{N}(\beta_{2,0} + \beta_{2,1} \times \text{elevation}_{\mathcal{A}_2} + \boldsymbol{w}_{\mathcal{A}_2}, \tau_2^2 \boldsymbol{I}),\\
\boldsymbol{I}_{\mathcal{S}_1} \mid \boldsymbol{\alpha}_1, \boldsymbol{\beta}_1, \boldsymbol{w}_{\mathcal{S}_1} &\sim \text{Bernoulli}\left(\text{logit}^{-1}(\alpha_{1,0} + \alpha_{1,1} \times (\beta_{1,0} + \beta_{1,1} \times \text{elevation}_{\mathcal{S}_1} + \boldsymbol{w}_{\mathcal{S}_1}))\right),\\
\boldsymbol{I}_{\mathcal{A}_2} \mid \boldsymbol{\alpha}_2, \boldsymbol{\beta}_2, \boldsymbol{w}_{\mathcal{A}_2} &\sim \text{Bernoulli}\left(\text{logit}^{-1}(\alpha_{2,0} + \alpha_{2,1} \times (\beta_{2,0} + \beta_{2,1} \times \text{elevation}_{\mathcal{A}_2} + \boldsymbol{w}_{\mathcal{A}_2}))\right).
\end{aligned}
\tag{11}
$$

We assume that both the rain gauge data and radar data are unbiased estimates of the underlying precipitation field, hence we set $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ and $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2$. Gaussian distributions are used for the log-transformed precipitation; hence, $\boldsymbol{w}_{\mathcal{A}_2}$ can be expressed as the averages of the latent process at sampling points within each area. In addition, we set $\tau_2^2 = \tau_1^2/5$ due to the averaging effect. The probability parameters for the Bernoulli distributions are transformations of the Gaussian means. A lower mean will lead to a lower probability of rain. In the model inference, we again use $m = 5$ nearest neighbours and $L = 5$ sampling points. We run four chains in parallel with 4,000 iterations for the fusion and fusion.vec models because they take longer to reach convergence and 2,000 iterations for the fusion.ncp model. In each case, half of the iterations are used as warm-up samples. In the model fitting, we set the precipitation to be zero if the probability parameter is less than 0.5. Although the model structure is slightly different from Section 2, it is straightforwardly implemented. The full model code is provided in the Supporting Information.

The goal in our case study is not to predict measurements but to obtain the underlying precipitation field; hence, we evaluate the goodness of fit for different implementations. Their results are summarized in Figure 2 and Table 2. As it can be seen in Figure 2, both implementations provided good fit to the observed precipitations. At locations where there was no observed precipitation, the models provided some nonzero estimates. Table 2 shows that the computation time for both fusion.vec and fusion.ncp is reduced, consistent with the simulation results.

6 | DISCUSSION

We have demonstrated the flexibility of the spatial fusion model framework in terms of tackling various change-of-support problems, which can occur due to observations measured at different spatial resolutions. Wang et al. (2018) focused their discussion on jointly modelling point and areal data. However, the proposed framework can be modified to handle analyses of point-point or area-area data fusion via the same solution to handling COSP. Many methods for tackling COSP are distribution-specific. The proposed spatial fusion model framework offers a distribution-free unifying approach. In addition, we have proposed an efficient implementation of fusion models, which takes advantage of vectorization in the Stan modelling language and noncentred parameterization. The new implementation improves the computational efficiency several fold. Such implementation can also be adapted in nonfusion spatial models that use NNGP.

The simulation study aimed to compare the efficiency of different model implementations. The noncentred parameterizations have shown consistently the best performance in all settings, including different spatial settings and different combinations of sample sizes. There are two main reasons for this increased computational speed. First, avoiding direct computation using NNGP likelihood but allowing a vectorized computation of Gaussian likelihoods avoided complex gradient evaluation. This shortened the computation time. Second, the latent process at observed and sampling locations are sampled as parameters in a Bayesian hierarchical model. This resulted in a high-dimensional parameter space with correlation between those latent process parameters and θ. Moving to a noncentred parameterization for both location and scale parameters improved the number of ESSs per minute. This implies faster convergence of the chains; hence, fewer numbers of iteration are required for model inference. In nonfusion Gaussian process models, the strength of spatial correlation among other parameters has an effect on the efficiency of different parameterizations (Bass & Sahu, 2017; Papaspiliopoulos, Roberts, & Sköld, 2003). However, it is beyond the scope of this study to investigate those effects in the spatial fusion model setting.

For spatial fusion models with a larger dataset, the spatial range parameter ϕ and/or partial-sill σ^2 can be fixed at some empirical estimates, for example, via maximum likelihood estimation, by using an expectation–maximization algorithm similar to that in Shi and Kang (2017). Under such an empirical Bayes approach, the covariance matrix structure becomes deterministic given the locations, hence saving a large amount of computation time during Markov chain Monte Carlo. An implementation of this for the case study dataset is shown in the Supporting Information. The performance improvement in terms of computational efficiency is about fivefold compared with the fusion.ncp implementation, at the cost of slightly increased RMSEs.

Existing spatial fusion models face two main challenges, namely, limited flexibility and slow computation. Several studies have shown that utilizing more than a single spatial data source can improve prediction performance (Cowles et al., 2009; Moraga et al., 2017; Shi & Kang, 2017; Wang et al., 2018). Therefore, having computationally efficient and flexible spatial fusion models enables more potential applications in areas such as geography, environmental science, and spatial epidemiology. Point process data are another form of available spatial data. They appear commonly in epidemiology, for example, in the location of disease occurrences. These data differ from point-reference data because their locations are random and not preselected. A framework for unifying three types of spatial data classified by Cressie (1991), namely, point-referenced data, point process data, and areal data, would be especially valuable for epidemiological applications. Finally, similar to other spatial models, a spatio-temporal extension is possible with the generalized spatial fusion model framework. In such an extension, an empirical Bayes approach is likely to be needed in order for the extension to be computationally feasible.

ACKNOWLEDGEMENTS

This work was supported by the Swiss National Science Foundation (Grant 175529). The precipitation data used in the case study were provided by MeteoSwiss, Federal Office of Meteorology and Climatology, Switzerland (https://www.meteoschweiz.admin.ch/home/service-und-publikationen/produkte.html).

ORCID

Craig Wang https://orcid.org/0000-0003-1804-2463
Reinhard Furrer https://orcid.org/0000-0002-6319-2332

REFERENCES

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*, pp. 136–139: CRC Press.

Banerjee, S., Gelfand, A. E., Finley, A. O., & Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(4), 825–848. https://doi.org/10.1111/j.1467-9868.2008.00663.x

Bass, M. R., & Sahu, S. K. (2017). A comparison of centring parameterisations of Gaussian process-based models for Bayesian computation using MCMC. *Statistics and Computing*, *27*(6), 1491–1512. https://doi.org/10.1007/s11222-016-9700-z

Berrocal, V. J., Gelfand, A. E., & Holland, D. M. (2010). A spatio-temporal downscaler for output from numerical models. *Journal of Agricultural, Biological, and Environmental Statistics*, *15*(2), 176–197. https://doi.org/10.1007/s13253-009-0004-z

Betancourt, M. J., & Girolami, M. (2013). Hamiltonian Monte Carlo for hierarchical models. arXiv:1312.0906 [stat.ME].

Bourgeois, A., Gaba, S., Munier-Jolain, N., Borgy, B., Monestiez, P., & Soubeyrand, S. (2012). Inferring weed spatial distribution from multi-type data. *Ecological Modelling*, *226*, 92–98. https://doi.org/10.1016/j.ecolmodel.2011.10.010

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455. https://doi.org/10.2307/1390675

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., ..., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Cowles, M. K., Yan, J., & Smith, B. (2009). Reparameterized and marginalized posterior and predictive sampling for complex Bayesian geostatistical models. *Journal of Computational and Graphical Statistics*, *18*(2), 262–282. https://doi.org/doi:10.1198/jcgs.2009.08012

Cressie, N. (1991). *Statistics for Spatial Data*. Hoboken, NJ, USA: John Wiley.

Cressie, N., & Johannesson, G. (2007). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(1), 209–226. https://doi.org/10.1111/j.1467-9868.2007.00633.x

Datta, A., Banerjee, S., Finley, A. O., & Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, *111*(514), 800–812. https://doi.org/10.1080/01621459.2015.1044091

Diggle, P. J., Tawn, J., & Moyeed, R. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *47*(3), 299–350. https://doi.org/10.1111/1467-9876.00113

Fuentes, M., & Raftery, A. E. (2005). Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. *Biometrics*, *61*(1), 36–45. https://doi.org/10.1111/j.0006-341X.2005.030821.x

Fuentes, M., Reich, B., & Lee, G. (2008). Spatial-temporal mesoscale modeling of rainfall intensity using gage and radar data. *The Annals of Applied Statistics*, *2*(4), 1148–1169. https://doi.org/http://www.jstor.org/stable/30245129

Furrer, R., Genton, M. G., & Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, *15*(3), 502–523. https://doi.org/10.1198/106186006X132178

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). New York: CRC Press.

Gotway, C. A., & Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, *97*(458), 632–648. https://doi.org/10.1198/016214502760047140

Greenland, S. (1992). Divergent biases in ecologic and individual-level studies. *Statistics in Medicine*, *11*(9), 1209–1223. https://doi.org/10.1002/sim.4780110907

Guan, Y., & Haran, M. (2018). A computationally efficient projection-based approach for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics*, *27*, 701–714. https://doi.org/10.1080/10618600.2018.1425625

Homan, M. D., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*(1), 1593–1623. http://dl.acm.org/citation.cfm?id=2627435.2638586

Lanckriet, G. R. G., De Bie, T., Cristianini, N., Jordan, M. I., & Noble, W. S. (2004). A statistical framework for genomic data fusion. *Bioinformatics*, *20*(16), 2626–2635. https://doi.org/doi.org/10.1093/bioinformatics/bth294

Liggins II, M., Hall, D., & Llinas, J. (2017). *Handbook of multisensor data fusion: Theory and practice*. Boca Raton, FL: CRC press.

Liu, Z., Le, N. D., & Zidek, J. V. (2011). An empirical assessment of Bayesian melding for mapping ozone pollution. *Environmetrics*, *22*(3), 340–353. https://doi.org/10.1002/env.1054

MeteoSwiss (2018). Weather radar network. https://www.meteoswiss.admin.ch/home/measurement-and-forecasting-systems/atmosphere/weather-radar-network.html Accessed: 2018-09-30.

Moraga, P., Cramb, S. M., Mengersen, K. L., & Pagano, M. (2017). A geostatistical model for combined analysis of point-level and area-level data using INLA and SPDE. *Spatial Statistics*, *21*, 27–41. https://doi.org/10.1016/j.spasta.2017.04.006

Nguyen, H., Cressie, N., & Braverman, A. (2012). Spatial statistical data fusion for remote sensing applications. *Journal of the American Statistical Association*, *107*(499), 1004–1018. https://doi.org/10.1080/01621459.2012.694717

Papaspiliopoulos, O., Roberts, G. O., & Sköld, M. (2003). Non-centered parameterisations for hierarchical models and data augmentation. In *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting* (Bernardo, J. M., Lindley, D. V., Dawid, A. P., Berger, J. O., West, M., Bayarri, M. J., & Heckerman, D., Eds.), Oxford University Press, USA.

Papaspiliopoulos, O., Roberts, G. O., & Sköld, M. (2007). A general framework for the parametrization of hierarchical models. *Statistical Science*, *22*(1), 59–73. https://doi.org/10.1214/088342307000000014

R Core Team (2018). *R: A language and Environment for Statistical Computing*, R Foundation for Statistical Computing. Vienna, Austria.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Sahu, S. K., Gelfand, A. E., & Holland, D. M. (2010). Fusing point and areal level space-time data with application to wet deposition. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *59*(1), 77–103. https://doi.org/10.1111/j.1467-9876.2009.00685.x

Shi, H., & Kang, E. L. (2017). Spatial data fusion for large non-Gaussian remote sensing datasets. *Stat*, *6*(1), 390–404. https://doi.org/10.1002/sta4.165

Sideris, I. V., Gabella, M., Erdin, R., & Germann, U. (2014). Real-time radar-rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland. *Quarterly Journal of the Royal Meteorological Society*, *140*(680), 1097–1111. https://doi.org/10.1002/qj.2188

Stan Development Team (2017). Stan modeling language users guide and reference manual. Version 2.17.0.

Stein, M. L. (2008). A modeling approach for large spatial datasets. *Journal of the Korean Statistical Society*, *37*(1), 3–10. https://doi.org/10.1016/j.jkss.2007.09.001

Wang, C., Puhan, M. A., & Furrer, R. (2018). Generalized spatial fusion model framework for joint analysis of point and areal data. *Spatial Statistics*, *23*, 72–90. https://doi.org/10.1016/j.spasta.2017.11.006

Zhang, L., Datta, A., & Banerjee, S. (2018). Practical Bayesian modeling and inference for massive spatial datasets on modest computing environments. arXiv:1802.00495 [stat.ME].

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Two zip files containing the data and the relevant code are available at www.math.uzh.ch/furrer/download/efficiency2018_stat_case.zip and www.math.uzh.ch/furrer/download/efficiency2018_stat_models.zip [Correction added on 28 January 2019, after first online publication: the links to the data files have been added.]

How to cite this article: Wang C, Furrer R. Efficient inference of generalized spatial fusion models with flexible specification. *Stat*. 2019;8:e216. https://doi.org/10.1002/sta4.216